DOCUMENT RESUME

ED 227 678 FL 013 540

AUTHOR Hamp-Lyons, Liz; Proulx, Greg
TITLE A Comparison of Two Methods of Teaching Advanced ESL Reading.
PUB DATE Jul 82
NOTE 22p.; Paper presented at the TESOL Summer Meeting (Evanston, IL, July 16-17, 1982). Appendix B was removed due to copyright restrictions.
PUB TYPE Speeches/Conference Papers (150) - Guides - Non-Classroom Use (055)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Cloze Procedure; Comparative Analysis; *English (Second Language); Higher Education; Instructional Materials; *Reading Instruction; Second Language Instruction; Second Language Learning; *Teaching Methods

ABSTRACT

Two methods of teaching advanced English as second language reading were compared in a sample of 24 Asian college students who were non-native speakers of English. The subjects were divided into three matched groups, each with one of two different teachers using different teaching methods, but both using the same text. It was hypothesized that all students would show progress when pretest and posttest scores were compared and that there would be no significant differences in posttest scores. The groups received either traditional or skills-based reading instruction (Group A, Teacher 1, traditional treatment; Group B, Teacher 1, skills-based treatment; and Group C, Teacher 2, traditional treatment). The hypothesis was partially supported; however, the group that received skills-based instruction scored significantly higher on the posttest and showed greater progress from pretest to posttest. The results show that a skills-based approach to teaching reading is more effective than a traditional approach. It is concluded that choice of teaching method does affect student learning even when the same instructional materials are used. (RW)

A Comparison of Two Methods of Teaching Advanced ESL Reading

Liz Hamp-Lyons[1]
Greg Proulx

WESL Institute
Western Illinois University

Abstract

Reading materials for ESL have recently undergone considerable changes, with the emphasis shifting from the traditional passage and exercises to an analysis of the discourse features of text in general. It is often assumed that "the right materials" will ensure student learning, and that poor student progress can be blamed on the textbook.

The purpose of this quasi-experimental investigation was to discover whether the teaching method used would have any significant effect on student learning when the same classroom materials were used, or whether the effect of the materials would be strong enough to make the treatment unimportant. Although the sample was small, and as with all classroom research, the variables could be neither counted nor accounted for, the investigation yielded some very interesting results.

Introduction

The investigation took place at WESL Institute, a seven level Intensive English Program attached to Western Illinois University under the auspices of International Programs. The data was collected in the spring of 1982 from all students enrolled at that time in the seventh (final) level of the program. The final level is a university preparation course for students

---

[1]Liz Hamp-Lyons is now at University of Edinburgh Institute for Applied Language Studies

scoring 500-550 on TOEFL, lasting for one semester (16 weeks) at ten hours per week, concurrent with a limited number of university credit courses. The course is divided into several modules, of which the academic reading module comprised 2 hours per week, ie. a total of 32 hours instruction in advanced, academic, reading. The book used for the course was Skillful Reading, by Amy L. Sonka (Prentice-Hall, 1981).

The subjects of the experiment were 24 non-native speakers of English, who had either scored between 500-550 on TOEFL, or successfully completed the previous level of our program. Nine of the subjects were graduates, while fifteen were preparing for entry to freshmen year. There were 15 males and 9 females. Native languages were:

Bahasa Malaysia	13
Korean	3
Chinese	3
Thai	2
Japanese	2
Vietnamese	1
	N = 24

It was hypothesized that all students would progress, ie. post-test scores would be higher than pre-test scores. It was further hypothesized that all groups would make the same or very similar mean improvement in score, regardless of the treatment they received. Intervening variables such as degree of acculturation, familiarity with the topics of the texts, etc. were assumed to operate equally within and among groups: thus, they were discounted as a measurable variable. It was hypothesized that the moderating variable, the teacher assigned to each group, would have no significant influence on post-test scores.

Pre/Post-Test Procedure

In this study, the twenty-four students were administered a cloze pre/post-test. This test consisted of six reading passages arranged in increasing order of length and difficulty. The passages varied from 350-500 words in length, and fifty blanks of equal length were inserted in each passage using fifth word deletion. The passages were selected from S.R.A. materials, after checking their levels using the FRY Readability Scale. The lengthier and more difficult passages contained extensive introductory and concluding material in order to provide the students with a fair amount of context.

The selection of the passages was carried out with the variety of cultures represented in the sample in mind. Neutral culturally oriented passages were selected in order that members of one culture group would not have an unfair advantage over another in their performance on the text.

During both the pre-test and the post-test, the students were allowed two hours to complete the test, without dictionaries or any other aids, and they were encouraged to guess whenever possible and to fill every blank.

There has been considerable research on the most valid method of scoring a cloze procedure. Recent studies have shown that the cloze is in no perceptible way less valid a measure when scored by exact word replacement than by semantic equivalent replacement. In this study, the pre/post-tests were scored by exact word replacement. Spelling errors were accepted if they were not meaning-reducing.

Forming the Groups

Since the subjects were to be divided into three groups, it was felt that given the variables under investigation, the size of the sample, and the various administrative constraints, a matched group design would yield the best results.

Firstly, five sets of three students were formed: raw scores on the pre-test were within six points of each other in each set. One student from each set was assigned to each of the three treatment groups. The remaining nine students who were "unmatchable" on the basis of their pre-test scores, were assigned to the three treatment groups, three to each group.

Although assignment of students to the treatment groups was random as far as possible, there were certain administrative constraints that prevented true random assignment. Initially, there was the aim to retain the total sample ratio of male to female within each group. Also, there was a desire to keep the languages and cultures as diverse as possible within each group. Finally, there was the advice of teachers who had taught the students in lower levels to keep certain students separated.

Using these criteria, the researchers formed three matched groups of eight students, each containing five males and three females. Group A was assigned to Teacher 1 for a traditional treatment; Group B was also assigned to Teacher 1, but received a skills-based treatment. Group C was assigned to Teacher 2 and received a traditional treatment.

In order to discover if the groups were evenly matched, ie. whether there were any significant differences among the means of the groups

on the pre-test, a Friedman Analysis of Variance (ANOVA) was run. Because of the wide variance of pre-test scores, the fact that there were only five matched sets and nine "unmatchables", and the aforementioned adminstrative constraints, an analysis of variance greater than .10 was determined to be significant. ANOVA revealed a difference among groups at the .149 level (see Figure 1).

A correlated t-test between the groups was run, and showed a significant difference between the means of group B ($\bar{x}$ = 83.9) and group C ($\bar{x}$ = 90.0) where $p \geq .05$. This may have been attributable to the placement of the "unmatchable" students, one of whom in the C group had an extremely high score. However, no significant difference was found between groups A and B, or between groups A and C. Although the mean for group C was higher than for the other two groups, this fact could not be altered, and the investigation went ahead.

The investigation aimed to test out the following hypotheses: firstly, that students will score significantly higher on a cloze post-test than on a cloze pre-test; and secondly, that among treatment groups, there will be no significant difference in scores on the cloze post-test.

Figure 1

Summary of Pre-Test Scores:

$\bar{X}_T = 86$ (All Students)

$S_T = 19.3$

$\bar{X}_A = 84.1$ Group A
$S_A = 16.7$ Tchr: 1
Method: Trad.

$\bar{X}_B = 83.9$ Group B
$S_B = 19.9$ Tchr: 1
Method: Skills

$\bar{X}_C = 90$ Group C
$S_C = 22.7$ Tchr: 2
Method: Trad.

ANOVA (Friedman) among groups:

$N = 8$ $X_r^2 = 3.93$ $p > .149$ (not sig.)

t-test between groups:

C - B
$t = 2.52$ $p > .05$

A - B
$t = .11$ not sig.

A - C
$t = 2.09$ not sig.

Design and Control of the Method Variable

The term 'skills-based' is used in this paper to refer to a range of approaches to the teaching of reading which may also be referred to as discourse analysis, which take their place under the broader umbrella of cognitive and communicative philosophies of language learning and teaching, and in which reading is taught as a number of sub-skills which can be integrated and generalized. There has been little quantitative research to find out whether in fact a discourse analysis approach to reading has had at least equal success in facilitating student progress as has the traditional approach, and no concrete evidence that it has been more successful in teaching foreign students to read in English, as its proponents imply or state.

Although we have spoken of skills-based versus traditional reading methods as though they were dichotomous, it would be more accurate to think of all methods as existing along one continuum, with each teacher operating at some point along that continuum, or more likely, at different points along it at different times and in different situations. That methods are in general difficult to pin down can be seen by a quick survey of a selection of the classical methodology texts (see Appendix A), and this seems to have been especially true for methods of teaching ESL reading. Therefore, although it was very easy to state the method variable as a dichotomy:

Method 1: Traditional (v) Method 2: Skills-Based

it was much more difficult to isolate the key features of each method and set up the research so that each group was only taught using genuinely typical techniques from one of the methods. The list below gives some

characteristics which were identified as typical of each method (note that most of them are based on experience and observation rather than upon literature of the method):

Traditional	Skills-Based/Discourse Analysis
product-centered	process-centered
reading is passive (receptive)	reading is active (interpretive; communicative)
bound to specific context	generalizable
focus on form	focus on meaning
schema not considered	schema important
reading as a form of behavior	affective domain considered
sentence level	discourse level
analytic	analytic↔synthetic
emphasis on content lexis	emphasis on structure lexis
discussion around text	discussion of text
emphasis on facts/details	emphasis on ideas/generalizations
one definition of 'comprehension'	comprehension varies with purpose/need
comprehension measured with questions: literal, inferential, critical	comprehension measured by varied activities: questions, logical manipulation eg. reorganization, outlining, non-linear response, etc.
glosses	no glosses
decoding→"right" answers	encoding→range of responses
reading aloud (teacher; student?)	silent reading
individual work	groups/pairs: sharing
teacher as authority	teacher as facilitator

It should be obvious that the two methods also have characteristics in common (for example, there is always a text and there are always questioning activities), and that in formulating a method variable and conducting an investigation into the effect of the method, we were quite aware that much of the time there would be no difference between activities. A large part of the treatment was constrained by the textbook, and in the main students did the same exercises from the book regardless of which treatment they were receiving (note: a discussion of the materials and how we controlled for treatment will follow). The method variable centered

on how the reading passages themselves were treated. In each treatment, there was always a session in which the main reading text of each chapter was intensively studied, and most of the methodological differences centered here, although clearly a different methodological base provided a different learning atmosphere for each group. To illustrate how two treatments of the same reading text might differ, two contrasting lesson plans, one for each method, are included (Figures 2 and 3):

Figure 2:

Treatment A: Traditional

Duration: 50 minutes
Preparation: read passage carefully at home (Chapter 2)

Lesson Plan:

take vocabulary questions (5 mins)
ask vocabulary questions (5 mins)
egs. prism; flashlight; filter
ask comprehension questions
a) literal (5 mins)
egs. Which wavelength bends the most?
When can you see a rainbow?
Why does grass look green?

b) inferential (10 mins)
egs. Why does a desert look yellow?
What colors does red reflect?
Why do windows have no color?

c) critical (10 mins)
egs. Why do yellow & blue light form white light?
What would happen if you shone a red light on a yellow object?
Why are the results of mixing light and paint different? Are they always different?

quiz: using vocabulary in context (10 minutes)

homework: outlining the passage

Figure 3:

Treatment B: Skills-Based

Duration: 50 minutes
Preparation: read the passage (Chapter 2); underline the topic sentence in each paragraph

Lesson Plan:

quiz: 10 mins
with the book shut, write the introduction in your own words; then do the same for the conclusion; list all the main points you can remember (10 mins)

analysis of the structure of text:

a) discussion of similarities and differences between introductions and conclusions (10 mins)

b) structure of the body (25 mins)
i) pairwork
ii) class discussion
iii) organization by levels of generality:

general → specific
specific → general

homework: go through the passage underlining all the example markers

Amy Sonka's Skillful Reading was seen as being suitable for the comparison of treatments because it is, relatively speaking, a middle of the road book. While the book identifies and treats a range of skills (eg. identifying topic sentences; scanning; skimming), it also treats reading traditionally with vocabulary activities and focus on grammatical structures. Its approach is more skills-based than traditional, but this was not seen as a problem since any reading passage can be treated traditionally.

We dealt with the problem of keeping the two methods separated in two ways. For the traditional group, we met every week to plan in detail the lessons for the following week; in our planning we agreed on the exercises from the book which would be used, constructed our own activities

to be used by both groups, fixed the amount of time to be spent on each activity, agreed on homework assignments, and generally reached a mutual strategy on every treatment question we were able to think of. In addition, in the first half of the semester we observed each other teaching our traditional groups once a week, and discussed our observations, with especial attention to any tendency by either of us to be tempted towards a discourse analytic treatment of any point. While it could not be claimed that these observations neutralized the effects of individual differences between teachers, they did allow insight into the nature and extent of those differences.

Secondly, in the two groups where there was one teacher (myself)[2] and two methods were used, I followed the lesson plan agreed on for the traditional group, and worked up a different plan for the skills-based group, which as can be seen from the example given earlier, de-emphasized the questioning based on the passage and instead worked intensively on discourse features (text structure; cohesion and coherence; anaphoric reference; logical connectors; etc.). It was not difficult to keep the methods separated because I was able to keep in mind what the other (traditional) groups were doing with the text and avoid those areas. I did not discuss the lesson plans for the skills-based group with the second teacher, in order to eliminate the possibility of skills-based activities overflowing into his treatment with his traditional group. Thus the second teacher acted as a control to show whether or not I was able to hold method as a true variable.

---

[2] 1st person singular refers to Liz Hamp-Lyons

Intervening Variables

As suggested earlier, the number of variables which could not be even partly controlled was as great as it always is in any research conducted in a naturalistic classroom setting as opposed to a laboratory setting. While the number of hours of tuition in reading and the period of time the course lasted was the same for all students, even such a simple factor as the amount of time spent on the homework was uncontrollable. Some other factors which undoubtedly functioned as intervening variables were: age (the sample included students aged between 18-27); whether or not they were taking university classes concurrently ; how many, and what type (most take 4-8 credit hours, but two students were not taking any courses); length of time in the U.S. (about half the students in the sample were new arrivals, while the other half had been in the country for at least four m..ths); major or intended major (the text we used was broadly science-based, and could be expected to be less motivating for a humanities-oriented student); whether motivation was integrative or instrumental (we have consistently found that instrumentally motivated students make most progress in our reading courses, whereas integratively motivated students do better in aural/oral skills). While some of the intervening variables can be identified, they cannot be measured, at least, not all at the same time, and cannot be taken into account when matching groups/sets. In addition, there certainly existed other intervening variables which have not even been identified. It can only be said that, without intending to deny or diminish their existence, we operated on the assumption that such factors would cancel each other out, ie. that in any group the range of influences and differences affecting the data in any direction would

be approximately equal. Such an assumption may or may not be reasonable, but it seems to be unavoidable in such ad-hoc classroom-based research as this.

Results

The first hypothesis, that students will score significantly higher on the post-test than on the pre-test, was tested statistically using the t-test for correlated samples.

The results of the t-test were highly significant (see Figure 4), indicating that there was a measurable difference between the sample's performance on the pre-test and the post-test. The negative t value supports the hypothesis that the students' scores on the post-test would be significantly higher than their scores on the pre-test, as had been predicted.

The second hypothesis, that there is no difference in student performance among treatment groups on the post-test was tested statistically using ANOVA. The results of the ANOVA showed a significant difference among groups (Figure 4), which warranted further analysis to determine exactly where the difference was.

The negative t-value in the first case showed that the performance of Group B was significantly higher than that of Group A. In the second case, Group B again showed a significantly higher performance than Group C. The negative t-value in the third case reveals a slightly higher performance for Group C than for Group A.

Further pre- and post-test analysis was done for each treatment group: while all three treatment groups show a significant difference

from pre- to post-test, Group B shows the highest t-value ($p > .001$, See Figure 4). This is especially interesting in view of the fact that group B's pre-test mean was noticeably lower than that of Group C.

When looking at the results, a memory effect was discounted as there was a sixteen-week interval between pre- and post-tests; practice effect was also discounted as a factor because the subjects did not have any other exposure to cloze procedure during the sixteen week interval.

Implications on the Findings

A comparison of the overall scores on the pre- and post-tests shows that the students did significantly better on the post-test: the sample mean rose from 86 to 109, ie. 23 points. The amount of teaching which the subjects received was approximately equivalent to one quarter of an academic year in high school: it was determined that a rise in score of 7.5 points approximated a rise in reading level of one quarter of a reading grade, ie. standard progress for the amount of teaching received. The mean rise in score of 23 points therefore represents a rise in reading level of approximately three-quarters of a reading grade. However, this comparison is a very rough rule of thumb, and no particular claims are being made. Our students were all literate in their first languages, and older than the equivalent high school student: it is felt that they should be expected to progress faster than high school students. It seems reasonable to state, however, that, regardless of which teacher or method they experienced, the subjects made greater improvement than they would have done without

Figure 4

Pre-Post Test t-test

t=-8.5 $p > .0005$ (one-tailed test)

Summary of Post-test Scores:

$\bar{X}_T = 109$
$S_T = 23.3$ (All Students)

$\bar{X}_A = 103$
$S_A = 17.1$ Group A, Tchr: 1, Method: Trad.

$\bar{X}_B = 115.25$
$S_B = 27$ Group B, Tchr: 1, Method: Skills

$\bar{X}_C = 110.8$
$S_C = 26$ Group C, Tchr: 2, Method: Trad.

ANOVA (Friedman) among groups:

N=8 $x_r^2 = 5.125$ $p > .10$ (sig.)

Two-tailed t-test between groups:

i) A - B
t = -2.91 $p > .05$

ii) B - C
t = 2.6 $p > .05$

iii) A - C
t = -1.99 $p > .10$

Pre-test - Post-test within groups; analysis on one-tailed t-test

Group A	t = 3.8	$p > .005$
B	t = 8.06	$p > .0005$
C	t = 6.4	$p > .0005$

any teaching.

The slight difference between groups A and C may be interpreted in various ways. The most likely cause is simply chance: however, it is possible that the second teacher allowed more skills-based activities to infiltrate his classroom that I did.

It can clearly be seen, on the raw data as easily as in the statistical analyses, that Group B, who received the skills-based treatment, had the highest mean score on the post-test, despite having had the lowest mean score on the pre-test. Group B's mean rose by 32 points, as compared to rises of 19 and 20 points for groups A and C respectively, both of which had received the traditional treatment. This difference is highly significant, and although teacher bias towards the skills-based treatment is admitted, we do not think that the bias was strong enough to account for the data. We believe that the results of the analysis show that a skills-based approach to teaching reading is more effective than a traditional approach. We are even more convinced of this in light of the fact that what we have referred to as the traditional approach used was not purely traditional: as stated earlier, the book teaches a number of reading skills, and practices them. In particular, the outlining activities are typically discourse-analytic, and thus our comparison was not of two extremes of method, but of only partial differences. A comparison of a true traditional approach with a fully skills-based one can be expected to reveal even more variation than was found here, and we hope that such a study will be done.

On studying the raw data it became clear that all students had not progressed equally; there was considerable deviation. This is inevitable

when such a limited sample is available; general intelligence, maturity, motivation, language learning ability, are all intervening variables which affect individual progress. It was, however, noticeable that the students who made the greatest improvements in scores were those from the middle range of the sample. The students who had been weakest and strongest on the pre-test in general made least progress. Discussing this finding, the teachers agreed that we had taught towards the middle of each group. In a sense it is gratifying to see that those to whom the teaching was specifically directed did in fact receive the most benefit, but at the same time, it is sad to realize that, even in groups as small as N=8, there are always some students who are not perfectly placed by level, and that this does have a disadvantaging effect upon them. The implication of this finding seems to be towards more individualization of ESL reading instruction.

One question which this investigation cannot help to answer is whether the more measurably successful skills-based approach actually helps the students to become better readers than they would have been if taught traditionally, or whether the students reach the same level of reading ability as they would have attained by <u>any</u> method (or even no method), but reach it sooner. To answer this question a longitudinal study would be needed, to follow the whole sample through at least a year of post-treatment reading, and ascertain whether the ex-students who had received the traditional treatment eventually caught up with those who had received the skills-based treatment; or conversely, whether the students who had made the greatest improvement eventually slipped back to the same level as those with lower post-test scores.

Such a question is very important and results of a longitudinal study would be very valuable.

Further Research

It seems clear that the issues raised here deserve further research. Where a larger population exists and the opportunity for multiple sections of a reading class using two or three different instructors, and/or two or three different methods is available, a replication of this study would provide further data in this area. Also, with a larger population, and with resulting larger individual sample sizes, the feasibility of using random sampling instead of matching would enable the researcher to use parametric analysis of variance and the F-test, which are more powerful tests than the non-parametric Friedman and the correlated t-test used in this design.

In addition, it appears that in this field there has been little quantitative research into the effect of different teaching methods on learning in skill areas other than reading. With the advent of the micro-computer and the many available programs in statistics, sampling, etc., it is becoming feasible for the classroom teacher to investigate the effects of variables such as those studied in this research. In this way, insights into the effects of the uncontrollable variables encountered in classroom research may be obtained by simply doing much more research and replication of research. Classroom teachers can test and evaluate for themselves the effectiveness of new methods and techniques constantly appearing in ESL, instead of accepting or rejecting such new developments without concrete evidence that they are or are not suitable for their own particular set of circumstances. Such replication of research is not really replication, since no two classroom situations

are identical. A large number of small scale, quasi-experimental studies which all reached similar conclusions would enable ESL professionals to make more powerful claims about the effectiveness of a method, or the influence of the teacher, or any other heavily investigated factor than are at present justifiable. While we feel that larger scale studies with greater internal validity are needed, we also feel that small scale, quasi-experimental "ad-hoc" classroom studies such as ours have a contribution to make, in the direction of greater pedagogic professionalism and away from an overconcern with the theoretical aspects of methodology. For the professional classroom teacher, the question which really matters is, "Does it work?" Large scale, fully controlled studies require large amounts of time, tremendous financial resources, and a large number of 'captive' students: all three of these are hard to find. Professional classroom teachers must find other ways of answering that question, in their own specific situation, satisfactorily. This is what we have tried to do.

While we do not claim that this investigation has yielded results which are definitive, we do believe that it has provided some evidence that a choice of teaching method <u>can</u> make a difference to measurable student improvement. We hope that this finding will lead other researchers to investigate this issue, and encourage classroom teachers of reading comprehension to re-evaluate their classroom approach, method and techniques and consider trying some variations, while carefully monitoring their efficacy or otherwise.

Appendix A: Select Bibliography of Methodology Texts

Allen, E.D. & Valette, R.M. (1972) Modern Language Classroom Techniques. New York: Harcourt, Brace, Jovanovich.

Brown, H. Douglas. (1980) Principles of Language Learning and Teaching. Englewood Cliffs, N.J.: Prentice-Hall.

Finocchiaro, M. (1958) Teaching English as a Second Language. New York: Harper & Row.

Lado, R. (1964) Language Teaching: A Scientific Approach. New York: McGraw-Hill.

Paulston, C.B. & Bruder, M. Newton. (1976) Teaching English as a Second Language: Techniques and Procedures. Cambridge, Mass.: Winthrop.

Rivers, W. (1968) Teaching Foreign Language Skills. Chicago: University of Chicago Press.

Bibliograpny

Allen, J.P.B. & Widdowson, H. (1974), 'Teaching the communicative use of English' in C.J. Brumfit & K. Johnson (eds.), The Communicative Approach to Language Teaching. (1979) Oxford, England: Oxford University Press.

Gaies, S. (1981) 'The fruits of labor: implications of research on classroom second language learning' in R. Orem & J. Haskell (eds.), Selected Papers from the First Midwest TESOL Conference. (1981) Champaign, Illinois: TESOL.

Goodman, K.S. (1967) 'Reading: A psycholinguistic guessing game' in H. Singer & R.B. Ruddell (eds.), Theoretical Models and Processes of Reading. (1970) Newark, Delaware: International Reading Association.

Grellet, Francoise (1981) Developing Reading Skills. Cambridge, England: Cambridge University Press.

Hamayan, E.V. & Tucker, G.R. (1980) 'Language input in the bilingual classroom and its relationship to second language achievement'. TESOL Quarterly 14:4, 453-66.

Smith, Frank (1978) Understanding Reading. New York: Holt, Rinehart & Winston.

Sonka, Amy (1981) Skillful Reading. Englewood Cliffs, N.J.: Prentice-Hall.

S.R.A.: Science Research Associates, Inc. (1970) Dimensions: Countries and Cultures. Chicago, Illinois.

Widdowson, H.G. (1978) Teaching Language as Communication. Oxford, England: Oxford University Press.

_____ (1979) Explorations in Applied Linguistics. Oxford, England: Oxford University Press.